



Predicting and measuring customer lifetime values for apartment tenants

Jian Wang  · Murtaza Das · Amar Duggasani

Received: 3 June 2016 / Accepted: 4 November 2016 / Published online: 8 December 2016
© The Author(s) 2016. This article is available at SpringerLink with Open Access.

Abstract It is of great benefit for an apartment community to predict the customer lifetime value (CLV) for each tenant. This prediction can be used not only to identify the most profitable residents, but also to make better pricing decisions, especially when optimizing renewal rents for expiring leases. CLV has been studied extensively in many industries such as service and retail. However, to our knowledge, there is no literature specifically addressing the estimation of CLV for apartment tenants. In this study, we propose an approximate approach to predicting the lifetime length and value for apartment tenants as well as their renewal probabilities. The model was estimated and tested based on a real dataset from 68 apartment companies in the US. The resulting prediction accuracy was particularly satisfactory for the tenants who did not renew or only renewed once.

Keywords Customer lifetime value · Apartment · Real estate · Prediction · Leases

1 Introduction

Increasingly, companies are viewing customers in terms of their lifetime value. In the apartment industry, the customer lifetime value (CLV) for a tenant measures the *total* value of the tenant for an apartment. CLV calculates the lifetime value that a tenant contributes during the tenant's entire stay in the apartment. In addition, the lifetime length for the tenant's entire stay is an important complementary measurement to CLV. It is advantageous for an apartment community to predict the lifetime length and value for each tenant. This prediction can be used not only to identify the

J. Wang (✉) · M. Das · A. Duggasani
The Rainmaker Group, 4550 North Point Parkway, Suite 400, Alpharetta, GA, USA
E-Mail: jwang@LetItRain.com

most valuable tenants, but also to make better pricing decisions, especially when optimizing renewal rents for expiring leases.

Apartments are one of the most important categories of commercial real estate. In the rental business of real estate, tenancies are created when the landlord and the tenant enter into a rental or lease agreement that conveys a possessory interest in the real estate to the tenant (Crane 2016). There are basically two types of tenancy: commercial tenancy and residential tenancy. Commercial tenants and residential tenants share some characteristics, e. g., both of them are leasing some property for a use, but they differ in several aspects. For example, the purposes of commercial tenants leasing a building are usually associated with a business operation like a store or an office, while those of residential tenants are tied with a personal residency such as single- and multi-family housing. The rental square footage for a commercial tenant tends to be, on average, larger than that for a residential tenant resulting in different tenant values. Apartments and apartment communities hereafter are limited to be the residential tenancy for multi-family housing.

Apartment communities belong to the service industry, but they have their own characteristics (Wang 2008). For example, the duration of a tenant is continuous. Before a prospective tenant moves in to an apartment, the person signs a new lease which specifies the unit to be rented, the move-in date, the chosen lease term (i. e., the number of months to stay), the monthly rent for the chosen lease term, as well as other terms. Before the lease is about to expire, the tenant will be offered a set of renewal options consisting of different lease terms at varying rents. If the tenant chooses to continue his stay in the apartment, he will renew the lease by signing another lease from the lease terms and rents of the renewal offer; otherwise, he will move out of the apartment by the time when the current lease expires. This renewal process will be repeated until the tenant moves out. Tenants rarely move back to the same apartment after moving out. This is largely because they may have relocated to a different city, moved to another apartment community, bought a house, etc.

Because of this unique characteristic of continuous residency, the lifetime of a tenant for an apartment is defined as the continuous duration between the times when the tenant first moves in and when he finally moves out. During a lifetime, the tenant will sign one or more leases, where the first lease is called a new lease, and any subsequent leases are called renewal leases. Therefore, the lifetime length and value of the tenant are calculated as the total length of lease terms and the accrued rents, respectively, from all the leases that the tenant has signed during a lifetime. Since tenants are not all the same, their leases are not necessarily the same thus leading to different lifetime lengths and values.

In this study, we propose an approximate approach to predicting and measuring the lifetime length and value for apartment tenants as well as their renewal probabilities. This paper is organized as follows: Sect. 2 presents a review of related literature. Sect. 3 describes a dataset of actual leasing transactions. This dataset is divided into two sample datasets to estimate and test an approximate approach, which is proposed in Sect. 4. In Sect. 5, we provide an empirical estimation of the proposed approach, followed by measuring accuracies. Finally, Sect. 6 concludes this paper with suggestions for further research and improvement.

2 Literature review

There exist many variants of CLV definitions, but they have similar meanings. For instance, CLV for a customer is often defined as the present value of all *future* profits generated from the customer (Gupta and Lehmann 2003). Research on the prediction of CLV has been performed mainly for retail and service industries. Because CLVs are closely related to the consumer behaviors, various methods for estimating CLVs have been suggested across industries (Fader et al. 2005). In measuring CLV for a customer, a common approach is to estimate the present value of the net benefit to the firm from the customer over time (generally measured as the revenues from the customer minus the cost to the firm for maintaining the relationship with the customer). Typically, the cost to a firm is controlled by the firm, and therefore is more predictable than the other drivers of CLV. As a result, researchers generally focus on a customer's revenue stream as the benefit from the customer to the firm.

Different models for measuring CLV result in different estimates of the expectations of future customer purchase behavior. For example, some models consider discrete time intervals and assume that each customer spends a given amount (e.g., an average amount of expenditure in the data) during each interval of time. This information, along with some assumptions about the customer lifetime length, is used to estimate the lifetime value of each customer by a discounted cash-flow (DCF) method (Berger and Nasr 1998).

Research on CLV measurement has so far focused on specific contexts. This is necessary because the data available to a researcher or a firm might be different under different contexts. Two types of context are generally considered: non-contractual and contractual (Reinartz and Kumar 2000, 2003; Borle et al. 2008). A non-contractual context is one in which the firm does not observe customer defection and the relationship between customer purchase behavior and customer lifetime is not certain (Fader et al. 2005). A contractual context, on the other hand, is one in which customer defections are observed, and a longer customer lifetime implies a higher customer lifetime value (Thomas 2001). Our problem tends to fall in the contractual context. When a prospect moves in or a tenant extends his stay in an apartment, he will sign or renew a lease with the apartment community. When he decides to move out, he will inform the apartment community about his decision. However, the apartment community does not know a tenant's lifetime length and value completely before he has physically moved out. This is because at each point in the tenant's lifetime, the apartment community is not certain whether the tenant will default his lease by moving out earlier, how many more times the tenant will renew, what lease terms the tenant will select, and what rents the tenant is willing to pay. In this regard, the tenant's lifetime length and value are uncertain.

There is a wealth of publications related to real estate. However, to our knowledge, none of them specifically addresses the estimation of CLV for apartment tenants. As defined early, the CLV of a tenant results from the accrued rents from one or more leases that the tenant signs during his lifetime. How the rents are set will thus impact the resulting CLV. On the other hand, the aggregate amount of CLVs from all individual tenants constitutes the gross rental revenue that the apartment will receive. This amount should reflect some kind of values that the

apartment property possesses, e. g., the “highest and best value” which, according to Miller and Geltner (2004), is defined as the reasonably probable and legal use of vacant land or improved property, that is physically possible, appropriately supported, financially feasible, and which results in the highest value at the date of the appraisal. As an attempt to obtain some insights about how to approximate CLV, it is helpful to understand how the rents and values of an apartment are determined.

Rent setting is an important and fundamental decision that apartment owners and operators have to make frequently. In traditional apartment management, rents are often set with the objective of maximizing occupancy and return on investment. Rents are normally determined by such factors as physical characteristics of a property, its current vacancy rate and competitive position in the market place, along with managers’ prior experience (Wang 2008). Sirmans and Benjamin (1991) performed an extensive literature review about the setting of apartment rents. For example, Simans et al. (1989) examined the effects of various factors on rent, in which the factors that they studied include amenities and services like covered parking, modern kitchen, and maid service, occupancy restrictions such as no pets allowed, traffic congestion, proximity to work, access to public transportation, and so on. Pagliari and Webb (1996) built a regression model to set rental rates based on rent concessions and occupancy rates. In modern apartment management, rents are often set with the objective of maximizing total revenue. For example, apartment revenue management (RM) proposes to optimize the rents by optimally balancing demand and supply in the consideration of competitor rents such that total revenue will be maximized (Wang 2008).

In apartment valuation, on the other hand, appraisers, investors, tax assessors and other real estate market participants are mainly the ones who are interested in estimating the values of properties. For example, when making income or DCF calculation of an apartment, an appraiser may take into account the lifetime value of a lease as well as many other factors that influence the values. His objective is to establish the market value of a property that accounts for the most probable price that would be paid for the property under competitive condition (Adetiloye and Eke 2014). The valuation of real estate has been studied extensively. There exist various valuation models such as cost approach, income capitalization approach, hedonic models, and so on, for estimating different types of values. Adetiloye and Eke (2014) give a thorough review on the real estate valuation. For instance, an appraiser may use the cost approach to estimate the market value by systematically estimating the cost of production (Miller and Geltner 2004). For apartment properties of real estate, in particular, there are also many researches. Zietz (2003) summarizes the empirical and theoretical studies for multifamily housing. As an example, Bible and Grablowsky (1984) observe that multifamily housing complexes located within restorative zoning neighborhoods increase in value at a higher rate than comparable complexes in neighborhoods that are not subject to restorative zoning codes.

The valuation of real estate, in general, and apartments, in particular, is sometimes performed by applying financial theory. Real estate investments comprise the most significant component of real asset investments. It is argued that real estate and financial assets share similar characteristics (Damodaran 2012). For instance, the accrued rents of a lease contract and the return of a bond or a stock represent

the expected cash flows on a real estate and a financial asset. Their values are determined by the cash flows they generate, the uncertainty associated with these cash flows and the expected growth in the cash flows. Since research on financial instruments is generally far ahead real estate research, many real estate researchers have therefore applied general financial theory to the valuation of real estate. There is a stream of literature relating real estate to financial assets. For example, McConnell and Schallheim (1983) priced leases and lease options using Black Scholes option pricing techniques. Wendt and Wong (1965) compare the investment performance of common stocks and apartment houses. They compare the DCF of FHA-financed residential projects with 76 randomly selected industrial stocks. They observe that because of the special real estate tax advantages, after-tax returns on equity investments in apartment houses are twice that of stock returns, but after-tax rates of returns vary significantly over different periods of time and across different properties.

In rent setting and valuation, the variables involved are more “exogenous” in the sense that they are more generic and related to the characteristics and conditions of the property and market where tenants reside. To a large extent, these variables play an important role in estimating the average rent and average value of a tenant, and thus the corresponding average CLV. Because CLVs are specific to individual tenants, other variables that are more “endogenous” should be taken into account so that we can better understand the CLVs on a personal level. Endogenous variables are those that are more specific and related to the characteristics and behaviors of individual tenants such as life expectancy, purchasing power, creditworthiness, household change, product promotion, default risk, lease term, number of renewal times, and so on. For instance, wealthier tenants may have a higher CLV because they can better endure rent increases or periods of unemployment. A tenant may increase his CLV by moving to a larger unit of the same apartment community when his family is growing. In addition, his CLV could also increase because larger families move less frequently leading to higher renewal probabilities. Also, a higher default risk increases the odds of moving out earlier than anticipated thereby decreasing his expected CLV.

In an attempt to increase the explanatory power of a statistical model, one may try to include as many variables as desired. In practice, however, there are several issues by doing so. First, the resulting error of such a fitted model may increase disproportionately. Second, it is difficult, if not impossible, and expensive to access all desired data, particularly, the endogenous data about personal demographics. Third, it is not easy to interpret a model when too many variables are included. As a result, in this study, we focused our attention on a small subset of variables that are related to leases. Our intention is to emphasize the underlying methodology of the proposed approach.

3 The dataset

The dataset consists of 77,536 historical leases from 62,643 tenants with lease dates ranging from March 4, 1995 to April 22, 2015. It was provided by The Rainmaker

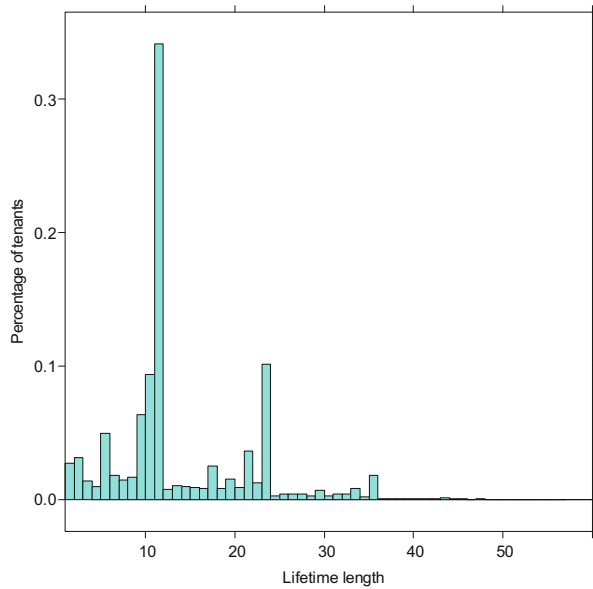
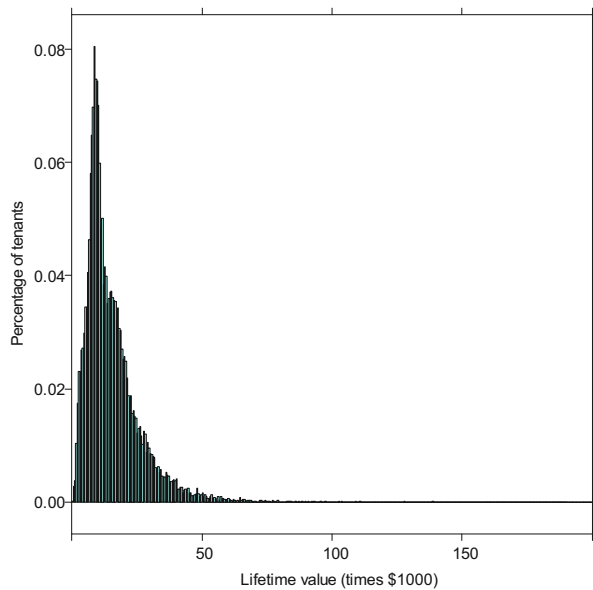
Group (www.letitrain.com), a revenue management software company providing pricing solutions for multi-family housing and gaming casino resorts industries. This anonymized dataset was selected from the historical transactions of 795 communities belonging to 68 apartment management companies across the United States. For each tenant in the dataset, we know his complete history of leasing transactions, i. e., total number of leases that he has signed, and the term, monthly rent, start date and end date of each signed lease. In this dataset, since there are more leases than tenants, we can infer that some tenants have signed two or more leases during their stays. In addition, this dataset also has 77,536 sets of 12 renewal options that were offered to the tenants when a lease was about to expire. Each renewal option specifies a rent for every lease term ranging from 1 to 12 months. A tenant has either chosen one of the 12 renewal options to renew, or he has chosen not to renew by moving out. From the dataset, we can compute the actual lifetime length and value for each tenant, which are equal to the total number of months, and the total amount of rents that he has paid during his stay. Furthermore, we can also compute the residual lifetime length and value for the remaining lifetime at each point in a lifetime. This dataset can thus be used to resemble a realistic context under which at each point in the lifetime of a tenant, we assume that the apartment community did not know how many more times the tenant would renew, what renewal lease terms the tenant would choose, and what monthly rents the tenant would pay. By doing so, we would be able to compare the predicted lifetime lengths and values with the actual ones.

Two samples are randomly drawn (without replacement) from this dataset. The first sample, referred to as the *estimation sample*, contains 62,049 leases from 50,181 tenants representing 80% of all the tenants. This sample will be used to estimate the parameters of a model to be proposed in Sect. 4. The second sample, referred to as the *validation sample*, includes 15,487 leases from 12,462 tenants representing the other 20% of the tenants. This sample will be used to predict and measure three dependent variables of primary interest: the residual lifetime length, the residual lifetime value, and the renewal probability, respectively, for each tenant during a lifetime.

Table 1 summarizes some descriptive statistics for the variables of lifetime length, lifetime value, number of renewals, and lease terms observed across all of the tenants in the estimation sample dataset. For example, on average, a tenant stays for about 14.3 months making a revenue contribution of \$ 15,352. The average number of renewals is 0.5 times with an average lease term 9.8 months.

Table 1 Summary Statistics

	Lifetime length (months)	Lifetime value (dollars)	Number of renewals (times)	Lease terms (months)
Mean	14.3	15,352	0.5	9.8
Std. Dev.	7.7	10,576	0.7	3.1
Median	12	12,375	1	12
Min	1	464	0	1
Max	48	138,655	5	12

Fig. 1 Tenants by Lifetime Length**Fig. 2** Tenants by Lifetime Value

Figs. 1, 2 and 3 display the histogram plots of lifetime length, lifetime value, and number of renewals, separately. They show significant heterogeneity across the tenants. Specifically, Fig. 1 shows the distribution of lifetime lengths ranging from 1 to 48 months. It can be seen that the peak occurs around 12 months representing 40% of tenants. This is consistent with the median of 12 months in Table 1, meaning that the majority of tenants sign a single lease with a lease term of 12 months.

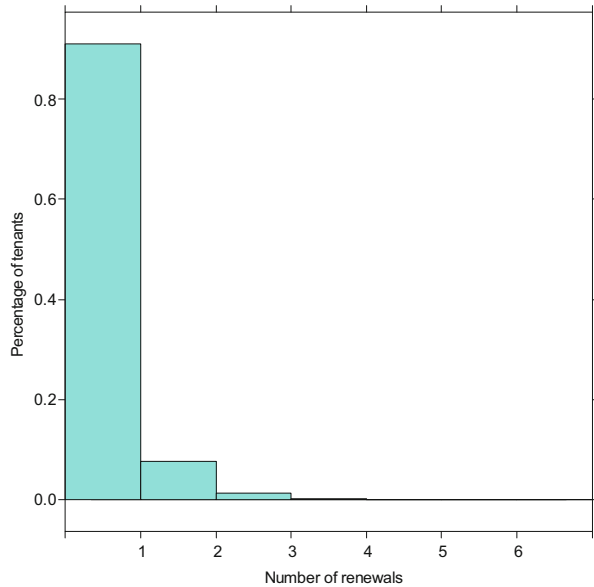
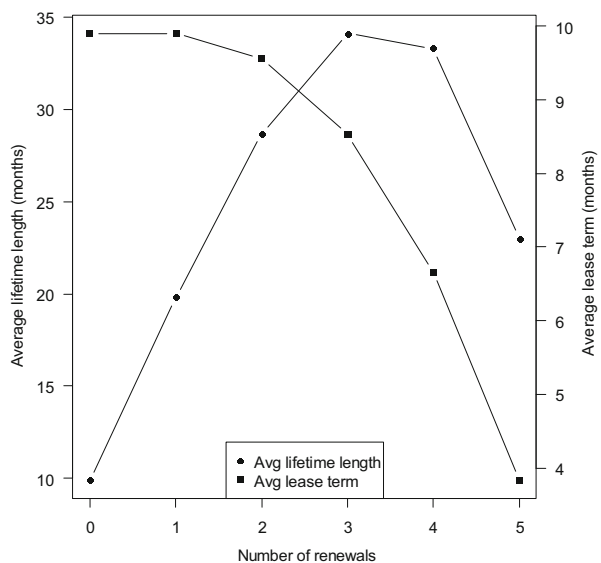
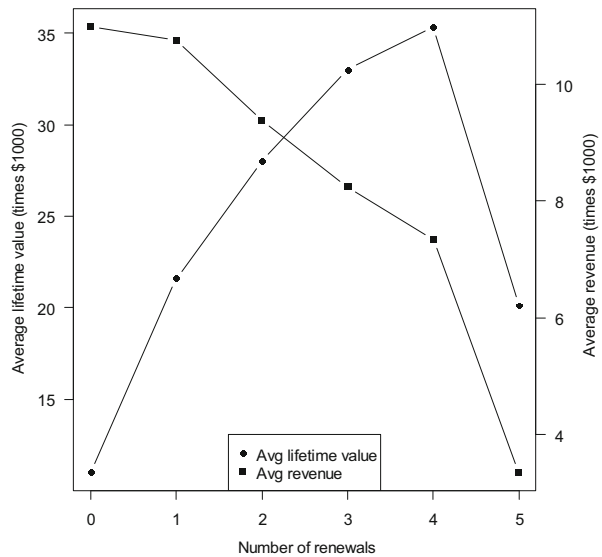
Fig. 3 Tenants by Number of Renewals**Fig. 4** Average Lifetime Length and Average Lease Term by Number of Renewals

Fig. 2 shows the distribution of lifetime values. In this Figure, although details are not shown, the lifetime values range from \$ 500 to \$ 138,600, among which more than 60% of tenants have a lifetime value of \$ 10430 or more. Fig. 3 shows that the number of renewals spans from zero to five times. The details have not been reported here, but more than 64% of tenants did not renew at all, around 26% of tenants renewed once, and the remaining 10% of tenants renewed two or more times.

Fig. 5 Average Lifetime Value and Average Revenue by Number of Renewals



The renewal behavior of tenants determines their lifetime lengths and values as well as the numbers of renewals. In this paper, two terminologies of *number of renewals* and *renewal times* are used throughout. *Number of renewals* is defined as the total number of renewal leases that a tenant has signed in a lifetime, while *renewal times* as the number of times that a tenant has been making a renewal decision at some point time in his lifetime. By this definition, total number of renewal times for a tenant is equal to the number of renewals plus one, because the tenant will not renew at the last renewal times.

Fig. 4 displays the average lifetime length and average lease term by number of renewals, where the number of renewals of zero is for the tenants who only signed a new lease and did not renew at all. It shows that the average lifetime length increases over number of renewals that is equal to or less than three. This seems intuitive because we would expect that the larger the number of renewals, the longer the lifetime length. However, the lifetime length starts to decrease over the number of renewals that is larger than three. This says that, on average, the lifetime length of a tenant does not necessarily become larger as the number of renewals increases. A plausible explanation is that tenants tend to sign shorter lease terms when they renew more times. The decreasing trend of average lease terms over number of renewals tends to support this explanation. Fig. 5 exhibits the average lifetime value and average revenue by the number of renewals. It shows that they maintain similar patterns as the average lifetime length and average lease term.

Fig. 6 illustrates the average residual lifetime lengths and values by renewal times. The residual lifetime length and revenue of a tenant at a renewal times are calculated as the sums of lease terms and revenues from the current and any future leases. As expected, both residual lifetime length and revenue decrease over renewal times. Because the maximum number of renewals is 5 in the estimation sample dataset, these values become zeros at the renewal times of 6.

Fig. 6 Average Residual Lifetime Length and Value by Renewal Times

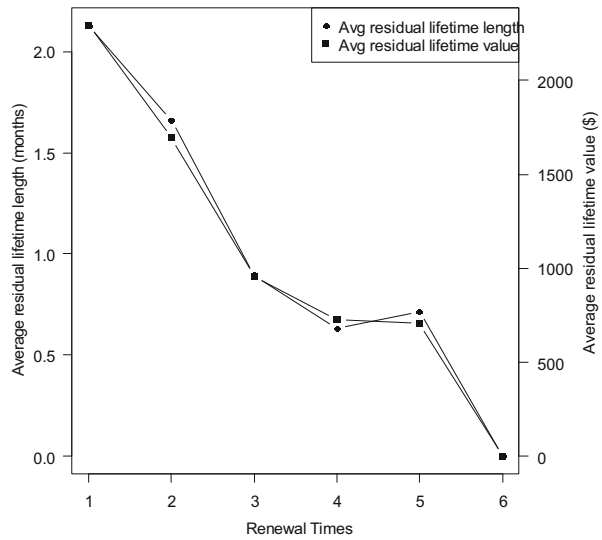
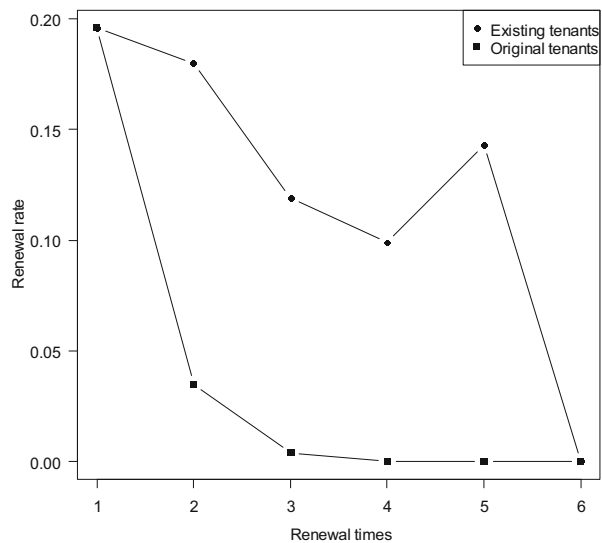


Fig. 7 Renewal Rates by Renewal Times



Finally, Fig. 7 displays two renewal rate curves by renewal times. The first renewal rate curve is defined as the fraction of the tenants who were active (i. e., still living in the apartment) at a given renewal times and had decided to renew. Specifically, at the first renewal times, about 20% of tenants chose to renew; at the second renewal times, about 18% of the remaining tenants chose to renew; and so on. By the last renewal times, all of the remaining tenants had moved out. On the other hand, the second renewal rate curve is defined as the fraction of the original tenants who chose to renew at a renewal times. It shows that the renewal rates of this curve decrease monotonically, meaning that the number of remaining active tenants

becomes smaller as renewal times increases, and that every tenant would move out by the last renewal times.

4 Model

The lifetime length and value, and renewal probability of an active tenant are determined by the renewal decisions of the tenant. When the current lease is about to expire, the tenant has two choices. He can choose either to renew for a particular lease term at an offered rent, or to move out. The renewal decision will impact the tenant's residual lifetime length and value. Fig. 8 illustrates a division of an active tenant's lifetime.

In Fig. 8, the lifetime of an active tenant is divided into three periods: past, current and future. A *past period* represents the time length of lease terms of all previous leases that the tenant has ever signed. A *current period* represents the time length of the lease term of the current lease. A *future period* represents the time length of lease terms of any future leases that the tenant may renew during the remainder of the tenant's lifetime. Furthermore, Fig. 8 illustrates a partition of the three periods into a number of consecutive segments, each of which contains a set of leasing options represented by arrows (directed edges) connected with circles (vertices). A dashed arrow represents a lease option that was offered or is to be offered, while a solid arrow represents an actual lease that the tenant has signed. Two circles at the ends of an arrow denote the start date and end date of a lease. Under this illustration, each segment in the past and current periods has one and only one solid arrow, while the segments in the future period have only dashed arrows.

Given any active tenant i , denote L_i and V_i as the tenant's lifetime length and value. At any time point t in the tenant's current period, L_i can be decomposed as $L_i = L_{i,p}(t) + L_{i,c}(t) + L_{i,f}(t)$, where $L_{i,p}(t)$, $L_{i,c}(t)$ and $L_{i,f}(t)$ represent the lifetime lengths for the past, current and future periods, respectively. Accordingly, V_i can also be decomposed as $V_i = V_{i,p}(t) + V_{i,c}(t) + V_{i,f}(t)$, where $V_{i,p}(t)$, $V_{i,c}(t)$ and $V_{i,f}(t)$ represent the lifetime values for the past, current and future periods at t . For both past and current periods, $L_{i,p}(t)$, $V_{i,p}(t)$, $L_{i,c}(t)$ and $V_{i,c}(t)$ are all known. They do not change over t , and can be calculated directly as the sums

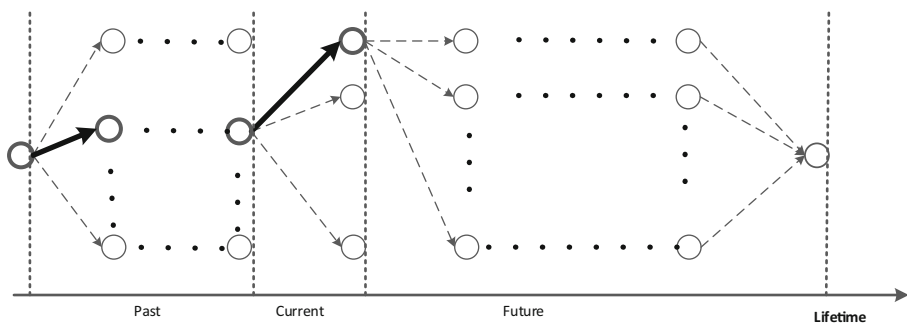


Fig. 8 Visual Depiction of the Division of a Lifetime

of lease terms and revenues (i. e., lease terms times monthly rents) from all the leases in the previous and current periods, respectively. For the future period, $L_{i,f}(t)$ and $V_{i,f}(t)$ represent the residual lifetime length and value from any future leases that might be signed after t . Both of them are unknown, and need to be estimated. For the sake of brevity, $L_{i,f}(t)$ and $V_{i,f}(t)$ will be called lifetime length and value at t in the remainder of the paper. We now propose an approach to approximating $L_{i,f}(t)$ and $V_{i,f}(t)$.

Denote τ as a discrete random variable corresponding to the renewal times at t , for which τ may take values of $1, \dots, N$, where N is the maximum renewal times to be allowed. Table 1 shows that the maximum number of renewals is 5 in our dataset, implying that $N = 6$ because every tenant had moved out by the sixth renewal times. Also, denote $\Pi(\tau) = (\pi_{l,j}(\tau))_{13 \times 13}$ as a renewal probability matrix, or called as transition matrix, at renewal times τ , where $\pi_{l,j}(\tau)$ is defined as the renewal probability from the current lease term l to the lease term j of a possible renewal lease, for $l, j = 0, 1, \dots, 12$. It is noted that when lease term j is equal to 0, it means that the tenant chooses to move out. When lease term l is equal to 0, there is no practical meaning. The reason that we introduce the notation of $l = 0$ here is just a matter of algebraic convenience. Specifically, for $\tau < N$, we set $\pi_{l,j}(\tau) \equiv 0$ for any j when $l = 0$. Also, we impose a condition that $\sum_{j=0}^{12} \pi_{l,j}(\tau) \equiv 1$ for any $l = 1, \dots, 12$, meaning that the tenant chooses either one of 12 lease terms to renew or chooses to move out. For $\tau = N$, we set $\pi_{l,j}(\tau) \equiv 0$ for any l and j , i. e., $\Pi(N) \equiv 0$, to comply with the assumption that N is the maximum renewal times. This condition implies that the estimates for $L_{i,f}(t)$ and $V_{i,f}(t)$ at $\tau = N$ are $\hat{L}_{i,f}(N) = \hat{V}_{i,f}(N) = 0$.

In addition, denote $\vec{l} = (1, \dots, 12, 0)$ as a 1×13 row vector representing all of the possible lease terms in which the special lease term of zero means a move-out. Suppose that l_i is the lease term of the current lease for tenant i . Let $\pi_{l_i,*}(\tau) = (\pi_{l_i,1}(\tau), \dots, \pi_{l_i,12}(\tau), \pi_{l_i,0}(\tau))$ be a 1×13 row vector of renewal probabilities of selecting the lease terms \vec{l} at τ by the tenant. This vector $\pi_{l_i,*}(\tau)$ is the l_i -th row from the renewal probability matrix $\Pi(\tau)$.

For any $\tau < N$, we propose to approximate $L_{i,f}(t)$ and $V_{i,f}(t)$ by

$$\hat{L}_{i,f}(\tau) = \pi_{l_i,*}(\tau) \left(1 + \sum_{s=\tau+1}^{N-1} \prod_{h=\tau+1}^s \Pi(h) \right) \vec{l}^T$$

And

$$\hat{V}_{i,f}(\tau) = \pi_{l_i,*}(\tau) \left(\overrightarrow{v_i(\tau)}^T + \sum_{s=\tau+1}^{N-1} \left(\prod_{h=\tau+1}^s \Pi(h) \right) \overrightarrow{v_i(s)}^T \right)$$

where $\overrightarrow{v_i(s)} = (v_{i,1}(s), \dots, v_{i,12}(s), 0)$ is a 1×13 row vector representing the revenues to be gained by choosing \vec{l} at renewal time s , for $s = \tau, \dots, N$. That is, $\overrightarrow{v_i(s)} = \vec{l} \circ r_i(s)$, a Hadamard product (or entry-wise product) of the lease

terms \vec{l} and the corresponding renewal rents $\overrightarrow{r_i(s)} = (r_{i,1}(s), \dots, r_{i,12}(s), 0)$. Specifically, $v_{i,j}(s) = j r_{i,j}(s)$, for $j = 1, \dots, 12$.

In the proposed formula, $\widehat{L}_{i,f}(\tau)$ and $\widehat{V}_{i,f}(\tau)$ are the sums of expected lease terms and expected revenues, from $(N - \tau)$ possible future renewal leases at renewal times $\tau, \tau + 1, \dots, N - 1$. Take a simple example to illustrate how $\widehat{L}_{i,f}(\tau)$ is estimated. Suppose that tenant i has a lease term of $l_i = 12$ for the current lease, It is the first time for the tenant to renew ($\tau = 1$), and the maximum number of renewal times allowed is 2 ($N = 2$). Then, $\widehat{L}_{i,f}(\tau)$ for $\tau = 1$ can be estimated as $\widehat{L}_{i,f}(1) = \pi_{12,*}(1) \vec{l}^T$, in which $\pi_{12,*}(1)$ and $\pi_{12,*}(1) \vec{l}^T$ represent the renewal probabilities and the expected lease term from of a possible future renewal lease. In a similar manner, $\widehat{V}_{i,f}(1) = \pi_{12,*}(1) \overrightarrow{v_i(1)}^T$ are the expected revenue from a possible future renewal lease.

It is noted that $\overrightarrow{v_i(s)}$ is derived from $\overrightarrow{r_i(s)} = (r_{i,1}(s), \dots, r_{i,12}(s), 0)$ and \vec{l} . For $s = \tau$, the renewal offer rents of $\overrightarrow{r_i(s)}$ usually become known within 30 to 60 days before the current lease expires. For $s > \tau$, $\overrightarrow{r_i(s)}$ are not known and need to be estimated. In the literature review, we mentioned that there exist a number of ways to estimate $\overrightarrow{r_i(s)}$. It is beyond the scope of this paper to discuss how to estimate $\overrightarrow{r_i(s)}$. To simplify our approximation, we will simply replace the values of $\overrightarrow{r_i(s)}$ for $s > \tau$ by the average renewal rents from the historical renewal offers.

Next, we describe the estimation of $\Pi(\tau)$. In practice, the determination of $\Pi(\tau)$ is complicated, and it is influenced by many factors including $\overrightarrow{r_i(s)}$. To simplify our estimation, we assume that $\Pi(\tau)$ satisfied the Markov property. Specifically, $\pi_{l,j}(\tau)$ are assumed to depend only the rents and lease terms of the current lease and a possible lease to be renewed. They are not dependent on the states of any previous leases that preceded the current lease. In this regard, for any renewal times $s > \tau$ for which the renewal options $\overrightarrow{r_i(s)}$ are not known, we will approximate $\pi_{l,j}(s)$ by using the empirical estimate of renewal probability; for the renewal times $s = \tau$ in which the renewal options $\overrightarrow{r_i(\tau)}$ are available, we will approximate $\pi_{l,j}(\tau)$ by using Multinomial Logit (MNL) model. MNL is a variant of customer choice models (Train 2009). The renewal probability $\pi_{l,j}(\tau)$ can be approximated by

$$\widehat{\pi}_{l,j}(\tau) = \frac{e^{V_{i,j}(l_i, r_i, r_{i,j}(\tau))}}{1 + \sum_{j'=1}^{12} e^{V_{i,j'}(l_i, r_i, r_{i,j'}(\tau))}}$$

where r_i denotes the rent of the current lease, and $V_{i,j}(l_i, r_i, r_{i,j}(\tau))$ a utility function (or called representative utility) at τ obtained by choosing the lease term j . $V_{i,j}(l_i, r_i, r_{i,j}(\tau))$ measures the perceived benefit of renewing a lease term j over choosing to move out. We assume that $V_{i,j}(l_i, r_i, r_{i,j}(\tau))$ is linear-in-parameter. Namely, $V_{i,j}(l_i, r_i, r_{i,j}(\tau)) = a_j(\tau) + \delta_{i,j}(\tau) b_j(\tau) + h_{i,j}(\tau) c_j(\tau)$, where the coefficients of $a_j(\tau)$, $b_j(\tau)$ and $c_j(\tau)$ are unknown parameters to be estimated, and $\delta_{i,j}(\tau)$ and $h_{i,j}(\tau)$ are alternative specific variables derived from l_i , r_i and

Table 2 Estimates of MNL Parameters for $\tau = 1$ and 2

	$\tau = 1$	$\tau = 2$		$\tau = 1$	$\tau = 2$
\hat{a}_1	-4.3* (0.17)	-4.6* (0.37)	\hat{b}_7	-3.6* (0.48)	-5.0* (0.99)
\hat{a}_2	-3.8* (0.11)	-3.8* (0.19)	\hat{b}_8	-4.5* (0.59)	-4.9* (1.25)
\hat{a}_3	-3.9* (0.10)	-3.7* (0.18)	\hat{b}_9	-4.2* (0.60)	-5.0* (1.31)
\hat{a}_4	-4.6* (0.14)	-4.4* (0.26)	\hat{b}_{10}	-2.4* (0.29)	-1.6* (0.60)
\hat{a}_5	-5.2* (0.19)	-4.8* (0.29)	\hat{b}_{11}	-4.2* (0.51)	-2.5* (1.23)
\hat{a}_6	-3.2* (0.04)	-3.2* (0.09)	\hat{b}_{12}	-3.1* (0.20)	-2.7* (0.49)
\hat{a}_7	-4.5* (0.07)	-4.8* (0.17)	\hat{c}_1	1.4* (0.42)	2.3* (0.48)
\hat{a}_8	-5.1* (0.09)	-5.2* (0.20)	\hat{c}_2	1.1* (0.16)	1.4* (0.24)
\hat{a}_9	-5.0* (0.09)	-5.4* (0.22)	\hat{c}_3	1.2* (0.14)	1.5* (0.23)
\hat{a}_{10}	-3.4* (0.03)	-3.3* (0.07)	\hat{c}_4	1.3* (0.27)	1.2* (0.49)
\hat{a}_{11}	-4.9* (0.07)	-4.9* (0.15)	\hat{c}_5	1.3* (0.43)	1.7* (0.74)
\hat{a}_{12}	-3.1* (0.04)	-3.3* (0.08)	\hat{c}_6	1.3* (0.07)	1.3* (0.12)
\hat{b}_1	-4.5* (0.57)	-2.9* (1.16)	\hat{c}_7	1.7* (0.17)	2.5* (0.30)
\hat{b}_2	-2.7* (0.42)	-2.3* (0.74)	\hat{c}_8	2.1* (0.22)	0.6 (1.02)
\hat{b}_3	-2.9* (0.41)	-2.9* (0.72)	\hat{c}_9	1.8* (0.23)	2.2* (0.55)
\hat{b}_4	-3.4* (0.59)	-2.8* (1.09)	\hat{c}_{10}	1.0* (0.07)	0.7* (0.14)
\hat{b}_5	-3.4* (0.80)	-4.6* (1.27)	\hat{c}_{11}	2.2* (0.10)	2.6* (0.23)
\hat{b}_6	-3.6* (0.25)	-3.7* (0.51)	\hat{c}_{12}	1.6* (0.04)	1.4* (0.10)

* indicates that the 95% posterior interval for a parameter does not contain 0

$r_{i,j}(\tau)$. Specifically, $\delta_{i,j}(\tau) = \left(\frac{r_{i,j}(\tau)}{r_i} - 1 \right)$ represents the relative rent change of the renewal rent $r_{i,j}(\tau)$ over the current rent r_i , and $h_{i,j}(\tau)$ is a habit formation (or inertia) describing whether a tenant tends to choose the same lease term as the current one when renewing, that is, if $j = l_i$, $h_{i,j}(\tau) = 1$; otherwise, $h_{i,j}(\tau) = 0$.

5 Estimation and validation

5.1 Empirical estimations

Renewal probabilities $\pi_{l_i,j}(\tau)$ were estimated based on the estimation sample dataset. As proposed early, when renewal rents $\overrightarrow{r_i(\tau)}$ are offered for some τ , $\pi_{l_i,j}(\tau)$ can be approximated via the estimation of MNL parameters; when $\overrightarrow{r_i(\tau)}$ are not available, $\pi_{l_i,j}(\tau)$ will be approximated empirically. In this estimation sample, more than 90% of tenants did not renew or just renewed once, which is not uncommon for apartments. This situation will cause the estimates of $\pi_{l_i,j}(\tau)$ to be inaccurate for $\tau \geq 3$ for which there does not have enough data. To alleviate this problem, we will approximate $\pi_{l_i,j}(\tau)$ using MNL model for $\tau = 1, 2$, and using the empirical estimates of probabilities for $\tau \geq 3$.

Table 2 reports the estimates of MNL parameters $\hat{a}_j(\tau)$, $\hat{b}_j(\tau)$ and $\hat{c}_j(\tau)$ for $j = 1, \dots, 12$ and $\tau = 1, 2$. The numbers in parentheses are the posterior standard deviations, and the superscript asterisks * indicate that the 95% posterior interval

Table 3 Empirical Estimates of $\Pi(\tau)$ for $\tau = 1$

	1	2	3	4	5	6	7	8	9	10	11	12	0
1	0.08	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.91
2	0.02	0.06	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.91
3	0.01	0.01	0.05	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.92
4	0.02	0.09	0.07	0.02	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.76
5	0.01	0.05	0.05	0.03	0.01	0.06	0.01	0.00	0.00	0.01	0.00	0.00	0.77
6	0.02	0.04	0.02	0.01	0.00	0.09	0.01	0.00	0.00	0.00	0.00	0.00	0.81
7	0.00	0.07	0.10	0.02	0.01	0.05	0.03	0.01	0.01	0.01	0.00	0.01	0.68
8	0.01	0.03	0.01	0.04	0.03	0.14	0.03	0.01	0.01	0.06	0.01	0.04	0.58
9	0.00	0.00	0.01	0.01	0.00	0.18	0.06	0.01	0.01	0.04	0.01	0.08	0.59
10	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.02	0.02	0.09	0.00	0.05	0.80
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.05	0.07	0.79
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.86

for a parameter does not contain 0. This is interpreted as an indicator of the estimate being statistically different from zero. It can be seen that the signs of $\hat{a}_j(\tau)$ are all negative implying that there was a larger tendency to move out than to renew. This implication is demonstrated in Fig. 7 also, in which, e. g., more than 80% of tenants chose to move out at the first renewal times. The signs of $\hat{b}_j(\tau)$ are negative as well, meaning that the larger the renewal rent change $\delta_{i,j}(\tau)$, the lesser the utility $V_{i,j}(l_i, r_i, r_{i,j}(\tau))$. This is consistent with intuitive expectation. Finally, the signs of $\hat{c}_j(\tau)$ are positive indicating that there exists a habitual inertia among tenants. Namely, when renewing, a tenant tends to renew to the same lease term as the current one.

As an illustration, Table 3 shows the empirical estimates of $\Pi(\tau) = (\pi_{l,j}(\tau))_{12 \times 13}$ for $\tau = 1$. The 12 rows represent the possible lease terms of a current lease, and the 13 columns the possible renewal lease terms, in which the renewal lease term of zero again represents the move-out option. On each row, the summation of the probability values across the 13 columns is equal to one, meaning that a tenant would either renew with one of 12 lease terms or choose to move out. It again shows that, as expected, the probabilities of moving out are larger than those of renewing. In addition, it can be observed that the diagonal entries of $\Pi(\tau)$ are not always larger than the off-diagonal entries. This empirical estimation does not provide strong evidence of the existence of habitual inertia as we saw in coefficient estimate of $\hat{c}_j(\tau)$ in MNL model.

5.2 Prediction and validation

We applied the proposed approach to predicting the lifetime length and value as well as renewal probability for each tenant in the validation sample, which consists of 12,462 past tenants with a total of 15,487 leases spreading across 4 renewal times. Because we already know the actual lifetime lengths and values and renewal outcomes of the tenants, we also tested the prediction performance of our approach.

Table 4 Comparison of actual and predicted lifetime length and value, and renewal probability

τ	1	2	3	4
$L_f(\tau)$	12.17	10.69	8.89	6.52
$\widehat{L}_f(\tau)$	11.84 (3%)	10.66 (0%)	9.56 (8%)	7.12 (9%)
$V_f(\tau)$	13,308.07	11,527.73	9190.05	7313.97
$\widehat{V}_f(\tau)$	13,147.28 (1%)	11,735.46 (2%)	9907.5 (8%)	8001.93 (9%)
$\pi_f(\tau)$	20%	18%	13%	7%
$\widehat{\pi}_f(\tau)$	20% (0%)	18% (0%)	17% (31%)	17% (143%)

For $\tau = 1, 2, 3, 4$, denote $L_f(\tau)$, $V_f(\tau)$ and $\pi_f(\tau)$ as the averages of the observed $L_{i,f}(\tau)$ and $V_{i,f}(\tau)$, and renewal probability from the validation sample, respectively. That is, $L_f(\tau)$ and $V_f(\tau)$ were calculated as the averages of actual lease terms and revenues from the current and subsequent leases with respect to τ . $\pi_f(\tau)$ was computed as the percentage of tenants who were active and had renewed at τ . Accordingly, denote $\widehat{L}_f(\tau)$, $\widehat{V}_f(\tau)$ and $\widehat{\pi}_f(\tau)$ as the averages of predicted $\widehat{L}_{i,f}(\tau)$, $\widehat{V}_{i,f}(\tau)$ and $\left(\sum_{j=1}^{12} \widehat{\pi}_{i,j}(\tau)\right)$, respectively. Table 4 summarizes the estimates of $L_f(\tau)$, $\widehat{L}_f(\tau)$, $V_f(\tau)$, $\widehat{V}_f(\tau)$, $\pi_f(\tau)$ and $\widehat{\pi}_f(\tau)$. The numbers in parentheses represent Mean Absolute Percentage Errors (MAPE) between the pairs of $L_f(\tau)$ and $\widehat{L}_f(\tau)$, $V_f(\tau)$ and $\widehat{V}_f(\tau)$, and $\pi_f(\tau)$ and $\widehat{\pi}_f(\tau)$, separately.

It can be seen that the prediction errors for $L_f(\tau)$, $V_f(\tau)$ and $\pi_f(\tau)$ were very small (i. e., MAPEs $\leq 3\%$) for $\tau = 1, 2$. However, for $\tau = 3, 4$, the prediction errors were not that satisfactory, particularly for $\pi_f(\tau)$. The main cause was because, as described earlier, the data was sparse (there are only less than 10% of tenants who had renewed two or more times). Therefore, this problem of data sparseness led to inaccurate estimates for $L_f(\tau)$, $V_f(\tau)$ and $\pi_f(\tau)$ for $\tau = 3, 4$.

6 Conclusion

In this study, we proposed an approximate approach to predicting the lifetime lengths and values for active tenants. We divided a sample dataset into estimation and validation samples. Based on the estimation sample dataset, we estimated the renewal probabilities. We then predicted the lifetime lengths and values as well as renewal probabilities for the tenants in the validation sample. The resulting prediction accuracy seemed to be satisfactory only for the tenants who did not renew or renewed once. It should be noted that in this article, lifetime lengths and values are forecast for the active tenants in the apartments of US market. As a consequence of the specifics of that market, the transferability of the results and applicability of the proposed model to other jurisdictions and cultures is limited.

To improve the prediction accuracy, the following explorations can be performed:

- *Inclusion of additional variables:* In this approach, we only use four variables: current lease term, renewal lease term, number of renewal times, and actual (or estimated) renewal rent offers. If accessible, we can consider more endogenous

and exogenous variables such as demographic information of age, income and family size, economic condition, market rents, migration tendency between states, and so on.

- *Utilization of existing data:* The prediction accuracy became unsatisfactory at the time when data amount was sparse, particularly for higher number of renewal times. To alleviate this issue, a set of hierarchy rules can be designed to pool the data of a lower number of renewal time for a higher number of renewal time. Although there is no rigorous academic proof about how much an improvement can be gained by doing so, this data pooling technique seems to be prevalent in practice.
- *Estimation of future renewal rent offers:* $\overrightarrow{r_i(\tau)}$ are unknown for any future renewal times τ , and need to be estimated. We estimated it with the average of renewal rent offers from historical expiring leases. As an alternative, we can consider to use other methods such as DCF and RM models.
- *Alternative customer choice models:* The MNL model that is used in estimating the renewal probabilities is popular in practice. It has many advantages such as being simple to understand, and easy to use. However, it sometimes suffers from an inherent assumption of Independence of Irrelevant Alternatives (IIA) (Meyer and Kahn 1991). This IIA assumption presumes that tenants would ignore the similarities among alternative lease terms when they make a renewal decision, which might not be always true. To mitigate this issue, other customer choice models such as Nested Logit (NL) model (McFadden 1981) can be taken into account. One of challenges of using NL model, for example, is to cluster “similar” lease terms into a group. Doing this “right” is not easy in practice. Some unsupervised learning techniques in data mining field might be needed.

The results may seem rudimentary, but they can still provide apartment communities with some insightful knowledge about the values of their tenants. A good estimate of CLV of the current tenants can be an additional key metric in assessing the financial value of the apartment property in comparison to other competing multifamily assets or other sister properties in an owner’s portfolio. Since the apartment industry has a competitive market environment, tenant behaviors might change quickly over time. As a consequence, the prediction of lifetime length and value cannot just be evaluated once and kept unchanged. They need to be updated regularly to reflect possible changes in tenant behaviors.

Acknowledgments We would also like to thank the three anonymous reviewer for their suggestions and comments. We would also like to show our gratitude to Ms. Marlene Rinker, our former colleague from The Rainmaker Group, for her revision on an earlier version of the manuscript, although any errors are our own and should not tarnish her esteemed reputation.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adetiloye KA, Eke PO (2014) A review of real estate and optimal pricing techniques. *Asian Econ Financial Rev* 4(12):1878–1893
- Berger PD, Nasr N (1998) Customer lifetime value: marketing models and applications. *J Interact Mark* 12:17–30
- Bible DS, Grablowsky BJ (1984) Restorative zoning effects on the valuation of multifamily income property. *Real Estate Apprais Anal* 51(2):32–36
- Borle S, Singh SS, Jain DC (2008) Customer lifetime value management. *Manage Sci* 54(1):100–112
- Crane F (2016) Real estate property management, 6th edn. Reality Publications, Riverside
- Damodaran A (2012) Investment valuation: tools and techniques for determining the value of any asset, 3rd edn. Wiley, Chichester
- Fader PS, Hardie BGS, Lee KL (2005) Counting your customers the easy way: an alternative to the Pareto/NBD model. *Mark Sci* 24(2):275–284
- Gupta S, Lehmann DR (2003) Customers as assets. *J Interact Mark* 17(1):9–24
- McConnell JJ, Schallheim JS (1983) Valuation of asset leasing contracts. *J Financ Econ* 12(2):237–261
- McFadden D (1981) Economic models of probabilistic choice. In: Manski CF, McFadden D (eds) *Structural analysis of discrete data with economic application*. MIT Press, Cambridge
- Meyer R, Kahn B (1991) Probabilistic models of consumer choice behavior. In: Robertson T, Kassarijan H (eds) *Handbook of customer behavior*. Prentice-Hall, Englewood Cliffs, pp 85–123
- Miller NG, Geltner DM (2004) Real estate principles for the new economy, 1st edn. SouthWestern Educational, Nashville
- Pagliari JL, Webb JR (1996) On setting apartment rental rates: a regression-based approach. *J Real Estate Res* 12(1):37–61
- Reinartz WJ, Kumar V (2000) On the profitability of long-life customers in a non-contractual setting: an empirical investigation and implications for marketing. *J Mark* 64:17–35
- Reinartz WJ, Kumar V (2003) The impact of customer relationship characteristics on profitable lifetime duration. *J Mark* 67:77–99
- Simans GS, Simans CF, Benjamin JD (1989) Determining apartment rent: the value of amenities, services, and external factors. *J Real Estate* 4(2):33–43
- Sirmans GS, Benjamin JD (1991) Determinants of market rent. *J Real Estate Res* 6(3):357–379
- Thomas JS (2001) A methodology for linking customer acquisition to customer retention. *J Mark Res* 38(2):262–268. doi:jmr.38.2.262.18848
- Train KE (2009) Discrete choice methods with simulations, 2nd edn. Cambridge University Press, Cambridge
- Wang J (2008) A realization of an apartment dynamic system. *J Revenue Pricing Manag* 7(3):256–265
- Wendt PF, Wong SN (1965) Investment performance: common stocks versus apartment houses. *J Finance* 20:633–646
- Zietz EM (2003) Multifamily housing: a review of theory and evidence. *J Real Estate Res* 25(2):185–244